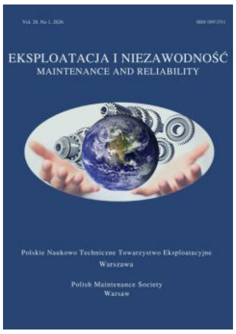

Eksploatacja i Niezawodnosc – Maintenance and Reliability

Volume 28 (2026), Issue 1

journal homepage: http://www.ein.org.pl

Article citation info:
Zhou W, Yang J, Chen J, Huang J, The Additive Reliability Model of Cyber Physical Systems Based on Local Polynomial Regression with Masked Failure Data, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2026: 28(1) http://doi.org/10.17531/ein/207794

The Additive Reliability Model of Cyber Physical Systems Based on Local Polynomial Regression with Masked Failure Data

Indexed by: Web of Science Group

Wenqian Zhou[a], Jianfeng Yang[b,c*], Jing Chen[b], Jiayue Huang[a]

[a] School of Mathematics and Statistics, Guizhou University, Guiyang 550025, China
[b] School of Mathematics and Statistics, Nanning Normal University, Nanning 530001, China
[c] Center for Applied Mathematics of Guangxi, Nanning Normal University, Nanning, 530001, China

Highlights

- Modeling the reliability of Cyber Physical System using local polynomial regression.
- Discussed the reliability research of Cyber Physical System under fully masked data.
- The masked data is allocated using multinomial distribution.

Abstract

In this paper, the cyber physical system is divided into a software system and a hardware system, and both the software system and the hardware system each contain several subsystems. To solve the difficult problem of parameter estimation in parametric reliability models such as the non-homogeneous Poisson process under masked data, this paper proposed an additive reliability model of cyber physical systems based on local polynomial regression under masked data. This model uses the multinomial distribution to allocate the masked data and employs the local weighted least squares method to solve the reliability model. Finally, by using a set of open-source software failure data and a set of simulation data in the cyber physical system, this paper conducts a performance comparison and analysis between the proposed non-parametric reliability model and traditional reliability models. The empirical results show that the proposed reliability model performs better in terms of the fitting effect and demonstrates stronger applicability and superiority.

Keywords

cyber physical system, masked data, local polynomial regression, additive reliability model

1. Introduction

In the current period of swift information technology advancement, the Cyber Physical System (CPS), as a key component of modern science and technology, is being increasingly widely applied in various fields. CPS realizes the in-depth integration of the physical world and the digital world through the deep fusion of communication, control and computing technologies. However, in practical applications, the reliability issues of CPS have become increasingly prominent and have become a key factor restricting its wide application. Especially in an environment with masked data, due to the complexity among the internal components of the system and the dynamic changes of the external environment, traditional reliability assessment methods find it difficult to accurately disclose the true situation of the system.

In the field of reliability engineering, accurately evaluating and predicting the reliability of a system is of crucial significance for ensuring its performance, improving safety and reducing costs. For example, power systems are confronted with strong uncertainties such as the intermittency of renewable energy sources, abrupt load changes, and equipment failures.

(*) Corresponding author.
E-mail addresses: W. Zhou (ORCID: 0009-0001-5225-7479) 2603222757@qq.com, J. Yang (ORCID: 0000-0003-3486-9604) yjf@nnnu.edu.cn, J. Chen 309656045@qq.com, J. Huang (ORCID: 0009-0004-4681-0495) 846489983@qq.com,

Traditional parametric models are prone to evaluation failures due to assumption biases. In contrast, nonparametric methods build models directly based on data, enabling more flexible handling of various data scenarios and the extraction of inherent reliability information from the data, thus providing a novel approach for reliability assessment. Xu et al.[1] proposed a new non-parametric density estimation method to derive and reconstruct the unknown distribution of the limit state function for reliability analysis. Li et al.[2] applied gene expression programming to non-parametric software reliability modeling. Barghout[3] put forward a new non-parametric model which considers the long-term trend and local behavior of reliability growth separately. The moving average method is used to capture the trend, while the local behavior is evaluated by the kernel estimator, and in this way, a software reliability model is established. When modeling software reliability, Wang et al.[4] et al. estimated the failure intensity function of the non-homogeneous Poisson process model through the constructed non-parametric method while taking into account the particularity of software failure data. Gweon et al.[5] et al. introduced a nearest neighbor-based recalibration method which aims to improve the reliability of probability classifiers in multi-class problems. Yu et al.[6] introduced a new method for adaptively reconstructing the unknown distribution of complex limit state functions and achieved structural reliability analysis through a non-parametric density estimation approach using harmonic transformation. Roy et al.[7] et al. developed a support vector regression model tailored for structural reliability analysis, they achieved this by addressing an optimization sub-problem, which aimed to reduce the mean square error value derived from the cross-validation technique to its minimum. Hu et al.[8] et al. proposed a kernel density estimation method, which is a non-parametric method for estimating the probability density function of wind speed, to conduct the reliability assessment of power generation systems.

The integrity and accuracy of data are crucial for constructing reliable reliability models. However, in practical situations, the situation of masked data is often encountered. The existence of such data has brought numerous challenges to the construction and analysis of reliability models, but it has also spurred many scholars to conduct in-depth research on reliability models under masked data. Cai et al.[9] estimated the reliability functions of systems and components under the usage stress level based on Type-I progressively hybrid censored data and masked data. Liu et al.[10] estimated the reliability of systems and components at specific time points based on the Lindley distribution with tampered failure rate model for masked system lifetime data. Sarhan et al.[11] calculated the reliability of system components through maximum likelihood estimation and Bayesian estimation based on masked system lifetime test data. In addition, Zhao et al.[12] proposed an additive non-homogeneous Poisson process model to describe the failure process of wireless sensor networks with subnets under masked data. Sarhan et al.[13, 14] considered the parameter estimation problem of component reliability in multi-component systems based on the situation where there are correlated masked system lifetime test data. Yang et al.[15] proposed a framework for an open-source software reliability growth model with masked data considering the failure detection and correction process, and a new expected least squares method was used to solve the parameter estimation problem. Zheng et al.[16], in order to evaluate the reliability of embedded systems more accurately, proposed an additive reliability model for software and hardware systems with masked data and failure correlation. Liu et al.[17] proposed a non-parametric Bayesian analysis method for correlated masked data under accelerated life tests with censoring and conducted reliability analysis. Kuo et al.[18] conducted Bayesian reliability modeling based on masked system lifetime data. Sarhan et al.[19] introduced how to use masked system lifetime data to estimate the reliability values of each component in a series system. Srivastava et al.[20] considered the analysis of masked data for different reliability systems (such as series, parallel and their various combined configurations). Wang et al.[21] proposed an approximate grey prediction method suitable for small samples and conducted reliability modeling based on masked system lifetime data. Sarhan et al.[22] estimated the parameters related to the lifetime distribution of system components using the maximum likelihood method based on masked data and also estimated the reliability of system components at specific times.

Due to the fact that CPS possess both the continuous dynamic characteristics of physical equipment and the discrete event-driven characteristics of information systems, and there is close interaction and collaboration between the two, this poses numerous challenges for constructing reliability models applicable to CPS. However, with the continuous expansion of CPS applications, research on its reliability models has increasingly become the focus of attention in both the academic and industrial communities. Alemayehu et al.[23] et al. used Markov chains to model and analyze the component reliability of CPS and proposed corresponding recovery techniques to ensure high reliability of the system and thus the continuity of system operation. Andrade et al.[24] et al. proposed a strategy based on stochastic Petri nets for the reliability modeling, evaluation, and adjustment of intelligent CPS. Lawal et al.[25] developed a meta-heuristic dynamic line rating sensor placement algorithm based on hierarchical clustering and proposed a framework for modeling the operational reliability of dynamic line ratings in cyber physical power systems networks. Yang et al.[26] introduced a reliability framework and assessment approach for CPS, incorporating communication failures, which was grounded in the instantaneous availability model and its analysis of fluctuations. Gholami et al.[27] considered the reliability modeling and evaluation of cyber physical power systems under the condition of multi-state independent components. He et al.[28] proposed a method for reliability modeling and evaluation of the cyberspace in cyber physical power systems to obtain the maximum flow that can meet the power demand. Liu et al.[29] modeled the reliability of active cyber physical distribution systems and established a reliability evaluation test system to analyze the impact of network failures. Gong et al.[30] studied a reliability model of a special cyber physical system with unreliable services and complex boundary behaviors. Rostami et al.[31] proposed an effective reliability evaluation technique for cyber physical power generation systems. Zhou et al.[32] studied the problem of optimizing the reliability of the embedded platform integrating central processing units and graphics processing units deployed in industrial cyber physical systems under temperature constraints. Ju et al.[33] analyzed the trade-off between reliability and security in cyber physical systems based on millimeter-wave ad hoc networks. Zhou et al.[34] analyzed the interlocking failures of the communication network in power cyber physical systems and proposed corresponding reliability evaluation methods. Huang et al.[35] introduced an innovative probabilistic modeling approach aimed at strengthening the analysis of the reliability of cyber physical systems affected by multilateral random attacks. Yang et al.[36] proposed using partially observable Markov decision processes to model cyber physical systems in an uncertain environment. Gong et al.[37] studied the reliability of a special cyber physical system with unreliable services and complex boundary behaviors.

To sum up, given the existence of masked data in systems, researchers have mostly considered establishing unified overall reliability models for systems, while relatively few have used additive models. Local polynomial regression is a non-parametric regression method that can flexibly capture the local features of data and is suitable for handling data with complex nonlinear relationships. Therefore, this paper proposed an additive reliability model of cyber physical systems based on local polynomial regression under masked data.

2. Related works

2.1. Masked data

During the process of system development and maintenance, system testing strategies and testing environments play a vital role. They not only affect the performance and functional verification of the system but also are directly related to the integrity of the system failure data records. In practical operations, due to the imperfection of testing strategies or the limitations of testing environments, a masking phenomenon may occur when obtaining failure data. This masking phenomenon means that although the system has failed, the exact cause of this failure remains unknown. Specifically, the information about the specific components (or subsystems) that caused the system failure is missing or cannot be accurately obtained. Among them, masked data[38] is a complex and challenging issue.

Let S_j be the set of components that cause the system to fail within time period $t_{j-1} - t_j$, where $S_j \subseteq \{1,2,\dots,k\}(j = 1,2,\dots,m)$. If any component of the system can cause the system to fail, then the data is masked data, that is, $S_j = \{1,2,\dots,k\}$. Let $t_j(j = 1,2,\dots,m)$ be the continuous observation time, and then there is the masked failure data of the Cyber Physical Systems (CPS), as shown in Table 1.

Table 1. Masked data.

Time	Component 1	Component 2	…	Component k	Masked data	Total	S_j
t_1	n_{11}	n_{21}	…	n_{k1}	n_{M1}	n_1	$\{1,2,\dots,k\}$
t_2	n_{12}	n_{22}	…	n_{k2}	n_{M2}	n_2	$\{1,2,\dots,k\}$
…	…	…	…	…	…	…	…
t_m	n_{1m}	n_{2m}	…	n_{km}	n_{Mm}	n_m	$\{1,2,\dots,k\}$

Among them, n_{ij} represents the number of failures detected for Component i of the CPS within the time period $t_{j-1}-t_j$, n_{Mj} represents the number of masked failures detected for the CPS within the time period $t_{j-1}-t_j$, n_j represents the total number of failures detected for the CPS within the time period $t_{j-1}-t_j$, the relationship among the three satisfies $n_j = n_{Mj} + \sum_{i=1}^{k} n_{ij}$. $i = 1,2,\dots,k$, $j = 1,2,\dots,m$, $S_j = \{1,2,\dots,k\}$.

2.2. Software reliability growth model based on non-homogeneous Poisson process

There are numerous types of CPS reliability models. Among them, for the software subsystems of CPS, the commonly applied ones are mostly reliability models of the non-homogeneous Poisson process (NHPP) type[39].

The assumptions of the NHPP-based software reliability models are as follows:

(1) The software failure process is a non-homogeneous Poisson process;

(2) The remaining defects in the software cause the current failures;

(3) The number of failures occurring in any time period is proportional to the number of remaining failures at that time;

(4) Perfect debugging.

Let $N(t)$ be the cumulative number of failures detected within time period $0, t$, and let $m(t)$ be the expectation of the cumulative number of failures at time t, that is, $E[N(t)] = m(t)$. Based on the above assumptions, we have:

$$\frac{dm(t)}{dt} = b(t)(a - m(t)) \quad (1)$$

Among them, a represents the total number of expected failures of the system, and $b(t)$ represents the failure detection rate function.

(1) When $b(t) = b$, the GO model is obtained, and its mean function is shown as follows:

$$m(t) = a(1 - e^{-bt}) \quad (2)$$

(2) When $b(t) = \frac{b^2 t}{1+bt}$, the DSS model is obtained, and its mean function is shown as follows:

$$m(t) = a(1 - (1 + bt)e^{-bt}) \quad (3)$$

2.3. Non-parametric regression model

Non-parametric regression[40] is a regression analysis method that does not impose too many restrictions on the distribution assumptions of data when conducting regression modeling. Unlike traditional parametric regression models (such as linear regression, logistic regression, etc.), non-parametric regression does not rely on functions in specific forms. Instead, it estimates the shape of the regression function based on the data itself. Since the distribution of the failure data of CPS is unknown and the CPS reliability model does not depend on a fixed functional form, this paper adopts the non-parametric regression model to model and fit the number of failures.

Suppose there is a set of cumulative failure data of CPS, denoted as $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$. If it is necessary to study the relationship between the cumulative failure number Y of CPS and time X, it can be expressed in the form of the following non-parametric regression model:

$$y_i = f(x_i) + \varepsilon_i, i = 1,2,\cdots,n \quad (4)$$

Among them, $f(\cdot)$ is the regression function, which is usually unknown, complex and nonlinear. The error term ε_i satisfies $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$.

3. Model construction

3.1. The additive reliability model of CPS based on non-homogeneous Poisson process

As shown in Figure 1, the CPS consists of only two subsystems, namely software subsystem S and hardware subsystem H. Software subsystem S contains r components, which are: $C_{s1}, C_{s2}, \dots, C_{sr}$. Hardware subsystem H contains l components, which are: $C_{h1}, C_{h2}, \dots, C_{hl}$. From the perspective of the sources of CPS failure data, a failure data may come from software subsystem S, or from hardware subsystem H, or may be jointly generated by software and hardware, that is, from the intersection part M of S and H. The software-hardware interface M contains n failure data, which are: $C_{m1}, C_{m2}, \ldots, C_{mn}$. If it is assumed that the failure data comes from M, then this failure data is masked data.

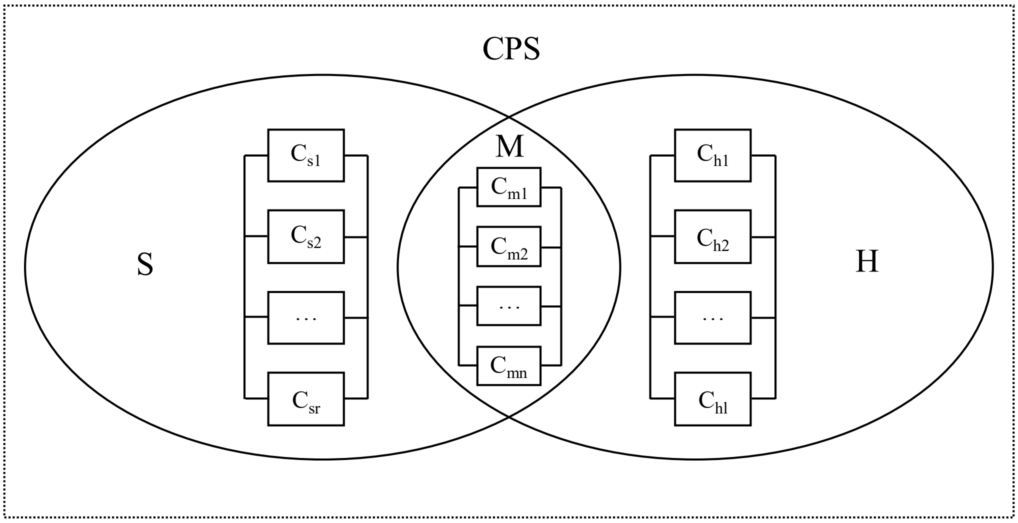

Figure 1. The structure diagram of failure data for CPS.

The reliability model of Cyber Physical Systems is a crucial tool for evaluating and optimizing system performance. Its reliability model aims to describe the logical relationships among various components of the system through mathematical methods and assess the reliability and availability of the system under given conditions.

Considering the situation where the software system of Cyber Physical Systems contains r component, the NHPP-based software additive reliability model based on masked data needs to add the following two assumptions on the basis of the software reliability model under ordinary failure data:

(1) The failure data $\{N_d(t), t \geq 0\}(d = 1,2, \ldots, r)$ of each component are mutually independent;

(2) The components that cause the system to fail may be any component of the system.

Based on the above assumptions, we have:

(1) When $b(t) = b$, the GO_SRM model is obtained, and its mean function is shown as follows:

$$m(t) = \sum_{d=1}^{r} a_d(1 - e^{-b_d t}) \quad (5)$$

(2) When $b(t) = \frac{b^2 t}{1+bt}$, the DSS_SRM model is obtained, and its mean function is shown as follows:

$$m(t) = \sum_{d=1}^{r} a_d(1 - (1 + b_d t)e^{-b_d t}) \quad (6)$$

For the reliability model of the hardware subsystem of CPS, the model used in Article[41] is adopted and denoted as the RH model, and its mean function is:

$$m(t) = \lambda t^{\beta} \quad (7)$$

Similarly, considering the situation where the hardware system of CPS contains l component, the failure data $\{N_h(t), t \geq 0\}(h = 1,2, \ldots, l)$ of each component are mutually independent, and $k = r + l$. An additive reliability model RH_SRM is established for the hardware system, and its mean function is:

$$m(t) = \sum_{h=1}^{l} \lambda_h t^{\beta_h} \quad (8)$$

In summary, the following NHPP-based additive reliability model for CPS is obtained:

(1) The GO model is adopted for the software subsystem, and the RH_SRM model is adopted for the hardware subsystem. This is denoted as the GO_CPS model:

$$m(t) = a(1 - e^{-bt}) + \sum_{h=1}^{l} \lambda_h t^{\beta_h} \quad (9)$$

(2) The DSS model is adopted for the software subsystem, and the RH_SRM model is adopted for the hardware subsystem. This is denoted as the DSS_CPS model:

$$m(t) = a(1 - (1 + bt)e^{-bt}) + \sum_{h=1}^{l} \lambda_h t^{\beta_h} \quad (10)$$

(3) The GO_SRM model is adopted for the software subsystem, and the RH_SRM model is adopted for the hardware subsystem. This is denoted as the GO_SRM_CPS model:

$$m(t) = \sum_{d=1}^{r} a_d(1 - e^{-b_d t}) + \sum_{h=1}^{l} \lambda_h t^{\beta_h} \quad (11)$$

(4) The DSS_SRM model is adopted for the software subsystem, and the RH_SRM model is adopted for the hardware subsystem. This is denoted as the DSS_SRM_CPS model:

$$m(t) = \sum_{d=1}^{r} a_d(1 - (1 + b_d t)e^{-b_d t}) + \sum_{h=1}^{l} \lambda_h t^{\beta_h} \quad (12)$$

In the NHPP model, the cumulative number of failures $N(t)$ follows a Poisson distribution, and its mean function $m(t)$ is of great significance. There is a relationship $R(t) = e^{-m(t)}$ between the reliability $R(t)$ and the mean function $m(t)$, indicating that the reliability is in the form of an exponential function of the mean function. The larger the value of the mean function is, the smaller the reliability will be. That is to say, as the average number of failures within a certain period of time increases, the probability that the product remains in normal operation at that time point decreases.

3.2. The additive reliability model of CPS based on local polynomial regression

The basic idea of local polynomial regression[42] is to use a polynomial to fit the data within the local neighborhood of each prediction point. Different from traditional polynomial regression, which fits a global polynomial to the entire dataset, local polynomial regression performs fitting in the local regions of the data, so that it can better capture the local characteristics of the data.

Suppose the dataset of the i-th subsystem of Cyber Physical Systems is $(x_i^j, y_i^j), i = 1,2,\ldots,k, j = 1,2,\ldots,m$. For a given prediction point x_i^0, the following polynomial function can be used to fit the local data points:

$$m_i(x_i^j) \approx \beta_i^0 + \beta_i^1(x_i^j - x_i^0) + \beta_i^2(x_i^j - x_i^0)^2 + \cdots + \beta_i^p(x_i^j - x_i^0)^p \quad (13)$$

Among them, $\beta_i^0, \beta_i^1, \cdots, \beta_i^p$ represents the coefficient of the polynomial, and p represents the order of the polynomial. The coefficients of the polynomial are estimated through the local weighted least squares method, that is, minimizing:

$$(\hat{\beta}_i^0, \hat{\beta}_i^1, \cdots, \hat{\beta}_i^p) = arg \min_{(\beta_i^0,\beta_i^1,\cdots,\beta_i^p)} \sum_{j=1}^{m} \left[y_i^j - \sum_{v=0}^{p} \beta_i^v (X_i^j - x_i^0)^v\right]^2 K_{h_i^m}(X_i^j - x_i^0) \quad (14)$$

Among them, $K_{h_i^n}(\cdot) = K(\cdot/h_i^n)/h_i^n$, $K(\cdot)$ is a Gaussian kernel function, which is used to measure the distance between data points and determines the weight of each data point, satisfying condition $\int_{-\infty}^{\infty} K(u)du = 1$. The kernel function should possess sufficient smoothness to ensure the continuity and differentiability of the estimation. The Gaussian kernel function selected in this paper has infinite-order smoothness, making it suitable for estimating cumulative failure data. $h_i^n > 0$ is the bandwidth, which controls the influence range of each data point of the i-th subsystem of Cyber Physical Systems in the estimation process. The choice of bandwidth parameter significantly impacts model performance. In this paper, the bandwidth parameter is determined through the cross-validation method. This method first divides the data into a training set and a validation set, then iterates over candidate bandwidths h_i^n, calculates the error on the validation set, and finally selects the h_i^n with the minimum error.

Define symbols:

$$X_i \equiv (x_i^1, x_i^2, \ldots, x_i^m)^T \quad (15)$$

$$Y_i \equiv (y_i^1, y_i^2, \ldots, y_i^m)^T \quad (16)$$

$$\beta_i \equiv (\beta_i^0, \beta_i^1, \cdots, \beta_i^p)^T \quad (17)$$

And:

$$W_i \equiv \begin{pmatrix} w_i^1(x_i^0) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_i^m(x_i^0) \end{pmatrix} \quad (18)$$

Then equation (16) can be written as:

$$\hat{\beta}_i = (X_i^T W_i X_i)^{-1} X_i^T W_i Y_i \quad (19)$$

Then, according to equation (15), we have:

$$\widehat{m}_i(x_i^0) = \hat{\beta}_i^0 = \ell^T \hat{\beta}_i = \ell^T (X_i^T W_i X_i)^{-1} X_i^T W_i Y_i \quad (20)$$

Among them, $\ell = (1,0,0,\ldots,0)^T$ is a $(p+1) \times 1$ dimensional vector. In practical applications, the order p is usually set to an odd number in order to reduce the corresponding boundary and design biases.

Summing up the cumulative number of failures of various subsystems in a CPS at the same time point and treating it as a whole, and then performing modeling based on the overall failure data at different time points through local polynomial regression, this model is denoted as the LPR model. This paper proposed the LPR_SRM model. That is, based on the two additional assumptions in Section 3.1, separate local polynomial regression modeling is conducted for each subsystem of CPS (including the hardware subsystem and the software subsystem). Finally, the local polynomial regression reliability models of each subsystem are added together to obtain the additive reliability model of CPS based on local polynomial regression under masked data.

3.3. The methods for solving models

3.3.1. Allocation of masked data

Suppose that for each component i, the cumulative number of failures observed at time t_j is m_{ij}, where $i = 1,2,\ldots,k$ and $j =$

$1,2,\ldots,m$. Due to the existence of masked data, m_{ij} is unknown. To solve this problem, let N_{ij} be the number of failures of component i within $t_{j-1},t_j)$, and the random vector $N_{*j}\sim M(n_j,p_{1j},p_{2j},\ldots,p_{kj})$, $N_{*j}=(N_{1j},N_{2j},\ldots,N_{kj})$, $n_j=\sum_{i=1}^{k}n_{ij}+n_{Mj}$. It can be known that:

$$p_{ij}=\frac{m_i(t_j)-m_i(t_j-1)}{m(t_j)-m(t_{j-1})}=\frac{m_i(t_j)-m_i(t_{j-1})}{\sum_{i=1}^{k}[m_i(t_j)-m_i(t_{j-1})]} \tag{21}$$

Obviously, $\sum_{i=1}^{k}p_{ij}=1$.

Therefore, the conditional probability is obtained:

$$P\left(\{\cap\langle N_{ij}\geq n_{ij}\rangle\}/\sum_{i=1}^{k}N_{ij}=n_j\right)=\sum_{\cap\langle r_i\geq n_{ij}\rangle,\sum_{i=1}^{k}N_{ij}=n_j}\left(n_j!\prod_{i=1}^{k}\frac{p_{ij}^{r_i}}{r_i!}\right) \tag{22}$$

Furthermore, the expected value of N_{ij} is calculated as follows:

$$\widehat{N}_{ij}=E(N_{ij}|n_{ij},n_{Mj})=\frac{\sum_{\cap\langle r_i\geq n_{ij}\rangle,\sum_{i=1}^{k}N_{ij}=n_j}\left(r_j!\prod_{i=1}^{k}\frac{p_{ij}^{r_i}}{r_i!}\right)}{\sum_{\cap\langle r_i\geq n_{ij}\rangle,\sum_{i=1}^{k}N_{ij}=n_j}\left(\prod_{i=1}^{k}\frac{p_{ij}^{r_i}}{r_i!}\right)} \tag{23}$$

After determining the estimated value $\widehat{N}_{ij}$ of m_{ij}, it will be used for subsequent estimations.

3.3.2. Locally Weighted Least Squares

Locally Weighted Least Squares[43] (LWLS) is a non-parametric learning method designed to address the limitations of traditional linear regression when handling nonlinear or locally varying data. Unlike ordinary least squares, LWLS assigns higher weights to samples near the target prediction point and lower weights to samples farther away, achieving a "local fitting" effect. Its core idea is to construct a locally approximated model near the target point rather than a unified global model.

LWLS estimates parameters by minimizing the weighted sum of squared residuals:

$$\hat{\beta}=arg\min_{\beta}\sum_{i=1}^{n}\omega_i(y_i-X_i\beta)^2 \tag{24}$$

ω_i is the weight of the i-th sample, X_i and y_i are the independent variable and dependent variable respectively, and β is the parameter to be estimated.

The weight is calculated by the Gaussian kernel, which reflects the distance between the sample and the prediction point:

$$\omega_i=exp\left(-\frac{(x_i-x)^2}{2\tau^2}\right) \tag{25}$$

x_i is the training sample point, x is the current prediction point, and τ is the bandwidth parameter that controls the local range.

By solving the weighted normal equation, the parameter estimator is obtained:

$$\hat{\beta}=(X^TWX)^{-1}X^TWY \tag{26}$$

W is the diagonal weight matrix and Y is the dependent variable vector.

3.3.3. Expectation Least Squares algorithm

The core idea of the Expectation Least Squares (ELS) algorithm[44] is similar to that of the Expectation Maximization algorithm. This algorithm aims to deal with the least-squares problems involving masked data and is an optimization iterative algorithm. The whole process can be divided into two major steps:

E-step: Allocate masked data

$$\widehat{N}_{ij}=E(N_{ij}|n_{ij},n_{Mj},\theta) \tag{27}$$

L-step: Minimize the sum of squared residuals

$$min\,S(\theta)=\sum_{i=1}^{k}\sum_{j=1}^{m}(m_i(t_j)-m_{ij})^2 \tag{28}$$

The fundamental procedures of the ELS algorithm are outlined below:

(1) Provide the initial value $(\theta_1,\theta_2,\ldots,\theta_k)^{(0)}$ of the unknown parameters;

(2) The existence of masked data leads to the unknown nature of m_{ij}. Solve for its estimated value $\widehat{N}_{ij}$;

(3) Calculate the probability: $P\left(\{\cap\langle N_{ij}\geq n_{ij}\rangle\}|\sum_{i=1}^{k}N_{ij}=n_j\right)$;

(4) The estimated value of m_{ij} is the conditional expected value $\widehat{N}_{ij}$ of the number of failures of each component;

(5) The new predicted value $(\theta_1,\theta_2,\ldots,\theta_k)^{(1)}$ is obtained through the L-step;

(6) Replace $(\theta_1,\theta_2,\ldots,\theta_k)^{(0)}$ with $(\theta_1,\theta_2,\ldots,\theta_k)^{(1)}$, and repeat the above steps (2) - (5) until the stable estimated value $\hat{\theta}$ is obtained.

3.3.4. Least squares estimator

Least squares estimator[45] (LSE) is based on an optimization strategy. Suppose there are a series of failure data points of CPS. The system failure probabilities $F(t_i)$ are observed at different time points t_i, and the reliability $R=f(t,\theta)$, t represents time, and θ is the parameter vector to be estimated. For each observed data point $(t_i,F(t_i))$, the predicted value of the model is $R_i=$

$f(t_i,\theta)$. The core principle of the least squares estimator is as follows:

$$S(\theta)=\sum_{i=1}^{n}(F(t_i)-f(t_i,\theta))^2 \quad (29)$$

By finding the parameter θ values that minimize $S(\theta)$, the parameters of the reliability model that can best fit the observed data are determined.

4. Case studies

4.1. Data description

4.1.1. Dataset 1

Dataset 1 is sourced from the real open-source software Apache Tomcat. As an outstanding open-source web application server, the Tomcat server has been widely used in the field of Java Web application development due to its characteristics such as being open-source and free, cross-platform, high-performance, and highly scalable. The intelligent parking system is used to manage data such as parking space information and vehicle entry and exit records in a parking lot. The Tomcat server supports the operation of relevant applications, enabling car owners to query parking space information and reserve parking spaces through mobile applications. Meanwhile, parking lot managers can conduct operations such as monitoring the status of parking spaces and fee management. Therefore, the Tomcat server is a software system of CPS, and its failure data belongs to the software part of CPS failure data. The real fault data of Dataset 1 in this paper originates from the fault tracking system of Tomcat 5, with its website URL being https://bz.apache.org/bugzilla/query.cgi. The data is integrated on a monthly basis, and a total of 40 sets of data from September 2008 to December 2011 are extracted, as shown in Table 2.

Table 2. Real failure dataset of Tomcat 5.

Time	F1	F2	F3	Masked	Time	F1	F2	F3	Masked
1	4	2	3	3	21	0	0	1	0
2	5	4	3	1	22	3	3	0	1
3	5	1	1	2	23	1	1	0	0
4	2	0	1	0	24	1	5	3	1
5	2	2	1	4	25	2	1	0	0
6	2	5	2	1	26	1	1	0	0
7	3	3	1	0	27	2	1	0	0
8	1	1	0	3	28	1	1	0	1
9	0	0	1	0	29	2	0	1	0
10	3	3	0	1	30	1	0	1	1
11	1	1	0	0	31	0	0	1	1
12	1	5	3	1	32	2	1	0	0
13	2	0	1	0	33	1	0	0	2
14	1	1	0	0	34	0	0	0	2
15	2	1	0	0	35	3	0	0	0
16	1	1	0	1	36	1	0	0	0
17	3	2	2	1	37	0	0	0	0
18	2	5	2	1	38	0	0	1	0
19	3	3	1	0	39	0	1	0	0
20	1	1	0	3	40	1	0	2	0

The table shows the cumulative failure data containing masked data. Among them, F1 indicates the count of malfunctions in the Catalina component, F2 indicates the count of malfunctions in the Connector and Webapps components, F3 indicates the count of malfunctions in the Jasper, Servlet and Native components, and "Masked" indicates the count of malfunctions in the "Unknown" ones, that is, the masked data in the system.

4.1.2. Dataset 2

Given that it is currently very difficult to directly obtain the detailed failure data of the actual hardware subsystems, this situation poses a significant challenge to the in-depth analysis and verification of the reliability models of hardware systems. To overcome this problem, this paper adopts simulation algorithms to generate the relevant failure data, thereby numerically verifying and analyzing the constructed reliability models of CPS.

Assuming that the failure of the software subsystem reliability model follows a GO model, with its failure mean value function expressed in the form of (30), and the failure of the hardware subsystem reliability model follows a power-law model, with its failure mean value function expressed in the form of (31).

$$m(t) = a(1 - e^{-bt}), b(t) = b \quad (30)$$

$$m(t) = \lambda t^{\beta}, b(t) = \lambda t^{\beta-1}, k > 0, \beta > 1 \quad (31)$$

Among them, a represents the expected total number of faults in the software subsystem; $b(t)$ represents the number of faults detected in the system per unit time; λ is a constant whose value is related to the complexity of the equipment; and β is an exponent.

Assuming that the total number of masked faults at each time point in the CPS follows a Poisson distribution, the fault counts of the software subsystem and hardware subsystem are aggregated with the masked data to obtain a complete simulation data table.

Simulations were carried out on the failure data of Software Subsystem 1, Software Subsystem 2, Software Subsystem 3, Hardware Subsystem 1, Hardware Subsystem 2 and Hardware Subsystem 3 based on masked data. The simulation results are shown in Table 3.

Table 3. Failure simulation dataset.

Time	1S(F1)	2S(F2)	3S(F3)	1H(F4)	2H(F5)	3H(F6)	Masked
1	29	28	26	1	2	2	4
2	22	19	20	5	7	4	4
3	26	19	18	2	10	7	11
4	17	20	19	4	6	6	0
5	23	20	20	4	10	9	6
6	13	8	11	5	13	10	1
7	14	7	6	4	23	8	12
8	16	13	12	2	19	16	8
9	9	5	6	5	27	23	7
10	3	8	4	5	17	16	8
11	6	10	8	4	29	24	3
12	8	6	7	5	25	20	9
13	6	8	6	7	26	18	0
14	12	8	9	6	28	19	2
15	5	5	4	5	30	22	6
16	4	5	3	4	38	29	3
17	3	4	2	8	24	28	9
18	8	2	3	4	27	24	11
19	1	4	2	3	28	22	8
20	3	3	1	10	31	23	1
21	4	4	2	1	46	27	9
22	7	3	4	4	32	18	8
23	2	5	3	6	36	16	13
24	2	1	2	13	37	23	9
25	0	0	1	6	44	26	10
26	3	2	0	5	38	19	4
27	2	1	1	6	42	18	5
28	2	1	0	10	53	21	9
29	3	0	1	8	39	25	1
30	2	0	3	10	56	31	1
31	0	1	2	9	60	28	4
32	1	0	2	9	56	26	2
33	2	0	1	8	52	33	5
34	1	1	1	7	70	35	9
35	1	1	1	11	60	28	5
36	0	0	1	8	62	26	1
37	1	0	0	7	73	29	4
38	0	1	0	12	72	31	6
39	1	1	1	8	47	34	4
40	2	0	1	11	73	28	8

As shown in Table 3, the CPS simulated by Dataset 2 contains six subsystems. Among them, there are three software subsystems, namely 1S (F1), 2S (F2) and 3S (F3), and there are also three hardware subsystems, namely 1H (F4), 2H (F5) and 3H (F6), with a total of 40 sets of data.

4.2. Model evaluation criteria

This paper conducts a comparative analysis of the proposed model and uses MSE (Mean Squared Error), AIC (Akaike Information Criterion), and BIC (Bayesian Information Criterion) as the indicators for evaluating the performance of the model.

The MSE calculates the differences between the predicted values and the actual values, squares these differences so that the positive and negative errors will not cancel each other out, and then calculates the mean of these squared errors to obtain a comprehensive error measurement indicator. Suppose there is a set of actual values $y_1, y_2, \dots, y_n$, and the corresponding predicted values are $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$. Its calculation formula is as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{32}$$

The AIC is a criterion used for model selection. When multiple different models are used to fit the same set of data, the AIC value of each model can be calculated. The smaller the AIC value, the better the relative goodness-of-fit of the model. Its calculation formula is as follows:

$$AIC = 2k - 2\,ln(L) \tag{33}$$

Among them, k signifies the quantity of parameters that need to be estimated within the model. Generally speaking, the more parameters there are, the more complex the model will be and the more data models it can fit. The likelihood function L measures the probability of observing the existing data under the given model parameters. The log-likelihood function $ln(L)$ is the natural logarithm of the likelihood function. The larger its value is, the better the model fits the data.

The BIC is derived based on the Bayesian theoretical framework. Similar to the AIC, its aim is to achieve a balance between the model's capacity to fit the data and the model's complexity in order to select an optimal model. Given a set of candidate models, the smaller the BIC value, the better the relative goodness-of-fit of the model. Its calculation formula is as follows:

$$BIC = -2\,ln(L) + k\,ln(n) \tag{34}$$

Among them, k denotes the quantity of parameters to be estimated in the model, $ln(L)$ represents the value of the model's log-likelihood function, and n represents the quantity of samples.

4.3. Comparative analysis of experimental results

4.3.1. Experimental results of Dataset 1

Based on the real failure dataset of Tomcat 5 shown in Table 2, the LSE is used to estimate the parameters for the GO model and the DSS model, the expected least squares algorithm is adopted to estimate the parameters for the GO_SRM model and the DSS_SRM model, and the local weighted least squares method is employed for non-parametric estimation of the LPR model and the LPR_SRM model. Since both the LPR model and the LPR_SRM model are non-parametric regression models, there are no parameter estimation values. Eventually, the parameter estimation values and performance comparison results of six software reliability models based on masked data are obtained, as shown in Table 4. The table indicates that: (1) Among the individual models, the LPR model has the best fitting effect, followed by the GO model, and the DSS model has the worst fitting effect and has a relatively large difference from the previous two models; (2) Among the superimposed models, the LPR_SRM model has a better fitting effect than the GO_SRM model and the DSS_SRM model; (3) Among the six models, the LPR_SRM model proposed in this paper has the best performance, followed by the LPR model, and the DSS model has the worst performance. Moreover, the performance of the superimposed models is better than that of the individual models.

Table 4. Parameter estimate values and performance comparison results of Dataset 1.

Model	Parameter estimate		MSE	AIC	BIC
GO	a=243.4513	b=0.0372	11.6866	104.3379	109.4045
GO_SRM	a1=103.6486 a2=97.3782 a3=42.0703	b1=0.0362 b2=0.0337 b3=0.0511	9.0797	94.2417	99.3083
DSS	a=183.9541	b=0.1315	55.8708	166.9217	171.9883
DSS_SRM	a1=77.8313 a2=71.9619 a3=35.0534	b1=0.1268 b2=0.1206 b3=0.1555	52.6594	164.5538	169.6205
LPR	-	-	8.1055	89.7019	94.7686
LPR_SRM	-	-	6.4374	68.5696	75.5696

The fitting graph of the cumulative number of software failures for the six models is shown in Figure 2. It can be clearly seen from the figure that the fitting effect of the models after additive processing is better than that of the models without additive processing, indicating that the superimposed models help to improve the accuracy of the evaluation of system reliability. In the time period [0, 10], the estimated values of the DSS model and the DSS_SRM model deviate significantly from the actual observed values. The remaining four models can all fit the real data well, and the fitting curves are roughly the same.

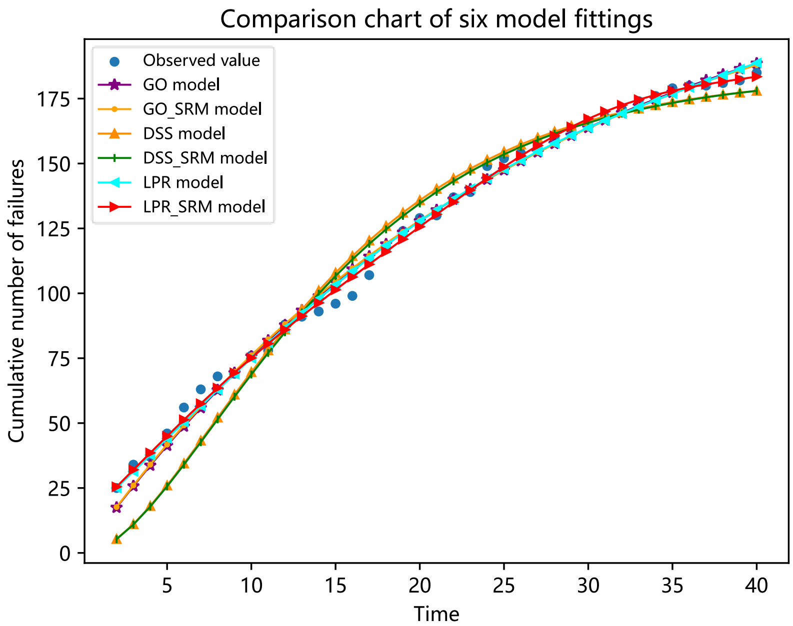

Figure 2. Fitting graph of the cumulative number of failures of the six models in Dataset 1.

4.3.2. Experimental results of Dataset 2

Based on the cumulative failure data of the system shown in Table 3, this paper first conducts parameter estimation for the models of the software subsystems of the CPS. The parameter estimation results of the GO model and the DSS model are obtained through the least squares estimator, and the parameter estimation results of the GO_SRM model and the DSS_SRM model are obtained through the expected least squares algorithm. The parameter estimation values and performance comparison results of the four software subsystem reliability models based on masked data are shown in Table 5. It is evident from the table that: (1) Among the individual models, the fitting effect of the GO model is better than that of the DSS model; (2) Among the superimposed models, the MSE of the GO_SRM model is much smaller than that of the DSS_SRM model, and the AIC and BIC of the GO_SRM model are also smaller than those of the DSS_SRM model, so the GO_SRM model is superior to the DSS_SRM model; (3) Overall, the fitting effect of the superimposed models is better than that of the individual models; (4) Among the four software subsystem reliability models, the MSE, AIC and BIC of the GO_SRM model are the smallest. Therefore, the GO_SRM model has the best fitting effect.

Table 5. Parameter estimation values and performance comparison results of software subsystem models.

Model	Parameter estimate		MSE	AIC	BIC
GO	a=807.6068	b=0.0988	314.2604	236.0089	241.0755
GO_SRM	a1=291.6684 a2=250.3145 a3=231.9635	b1= 0.0995 b2=0.1039 b3=0.1104	39.6814	153.2353	158.3019
DSS	a=729.6062	b=0.2509	830.3073	274.8718	279.9385
DSS_SRM	a1=274.1690 a2=237.0318 a3=218.5864	b1=0.2418 b2=0.2465 b3=0.2683	823.9320	274.5635	279.6302

The fitting graph of the cumulative number of failures of the four models is shown in Figure 3. It can be seen from the figure that after the additive processing is carried out on the GO model, the fitting effect of the new model obtained, namely the GO_SRM model, is better. Therefore, adding on the GO model helps to optimize and improve the performance of the model. However, after the additive processing is carried out on the DSS model, the fitting effect of the new model obtained, namely the DSS_SRM model, has not been significantly improved, and the fitting curve of the DSS_SRM model is roughly the same as that of the DSS model. Therefore, for the dataset in this paper, the additive processing carried out on the DSS model has a relatively small impact on the model performance. The reason is that when simulating this dataset, it was assumed that the failures of the software subsystem reliability model followed the GO model. Therefore, whether using the DSS model or the DSS_SRM model, neither achieved satisfactory fitting performance, and the difference between them was not significant. By comparing the fitting curves of the four models, It can be inferred that the GO_SRM model has the optimal fitting performance.

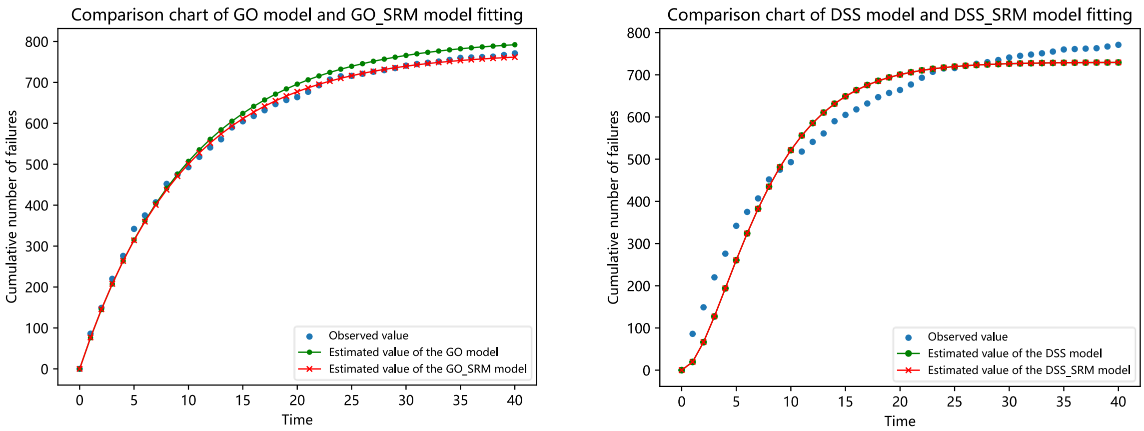

Figure 3. Fitting graph of the cumulative number of failures of the four models in Dataset 2.

Conduct parameter estimation for the models of the hardware subsystems of the CPS. The parameter estimation results of the RH_SRM model are obtained through the LSE. The parameter estimation values and evaluation indicator results of this model are shown in Table 6. Since the hardware subsystem contains three components, there are three sets of parameter estimation values for the RH_SRM model. The values of its MSE, AIC and BIC are all relatively small, indicating that the RH_SRM model has a relatively good fitting effect.

Table 6. Parameter estimation values and evaluation indicator results of the RH_SRM model.

Model	Parameter estimate		MSE	AIC	BIC
RH_SRM	λ_1=1.2354 λ_2=2.6151 λ_3=4.3913	β_1= 1.4617 β_2=1.7279 β_3=1.4476	105.5467	175.5852	180.6519

The fitting graph of the cumulative number of failures of the RH_SRM model is shown in Figure 4. Judging from the fitting curve and the observed value curve of this model, the two are very close, so it can be considered that the RH_SRM model has a relatively good fitting effect.

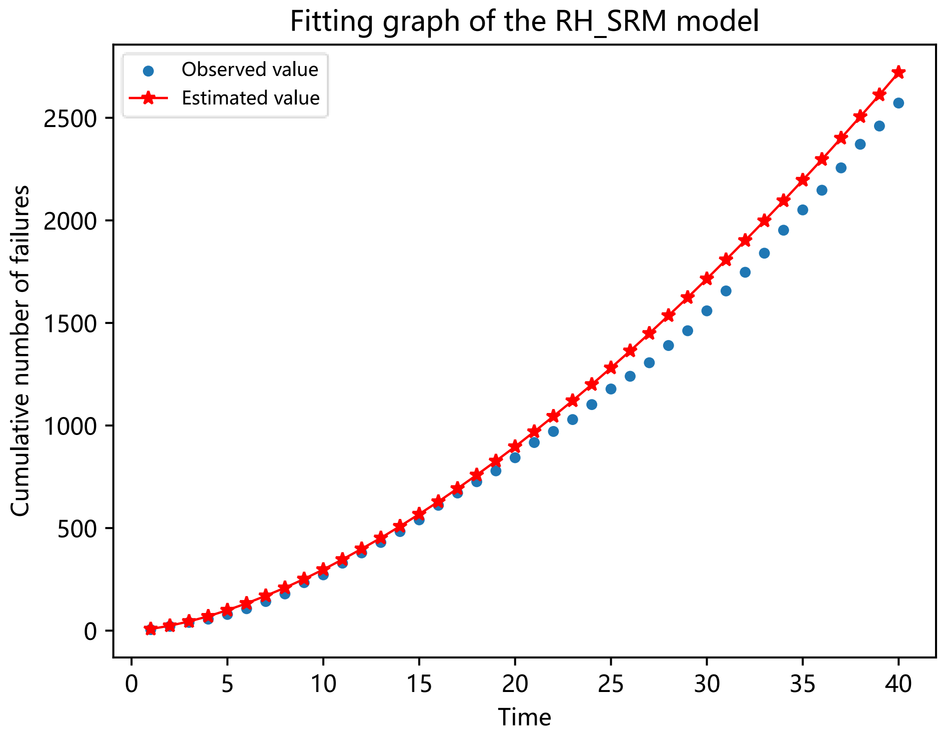

Figure 4. Fitting graph of the cumulative number of failures of the RH_SRM model.

Conduct performance comparison of the models for the CPS as a whole. Since both the LPR model and the LPR_SRM model proposed in this paper are non-parametric regression models, there are no parameter estimation values, and the results of their evaluation indicators are directly calculated. The evaluation indicator values of the GO_CPS model, the DSS_CPS model, the GO_SRM_CPS model and the DSS_SRM_CPS model are the mean values of the evaluation indicator values of the corresponding combined models. The results of the evaluation indicators of the above six models are shown in Table 7. By comparing the evaluation indicator values of the six models, it can be known that the MSE, the AIC and the BIC of the LPR_SRM model are the smallest, followed by the LPR model, then the GO_SRM_CPS model, and the DSS_CPS model has the worst fitting effect.

Table 7. Results of the evaluation indicators of the six models.

Model	MSE	AIC	BIC
LPR	58.5830	146.8178	150.8844
LPR_SRM	19.0131	115.2730	120.3397
GO_CPS	209.9035	205.7970	210.8637
GO_SRM_CPS	72.6140	164.4102	169.4769
DSS_CPS	467.9269	225.2285	230.2952
DSS_SRM_CPS	464.7393	225.0743	230.1416

The comparison of the fitting graphs of the LPR model, the LPR_SRM model, the GO_SRM_CPS model and the DSS_SRM_CPS model is shown in Figure 5.

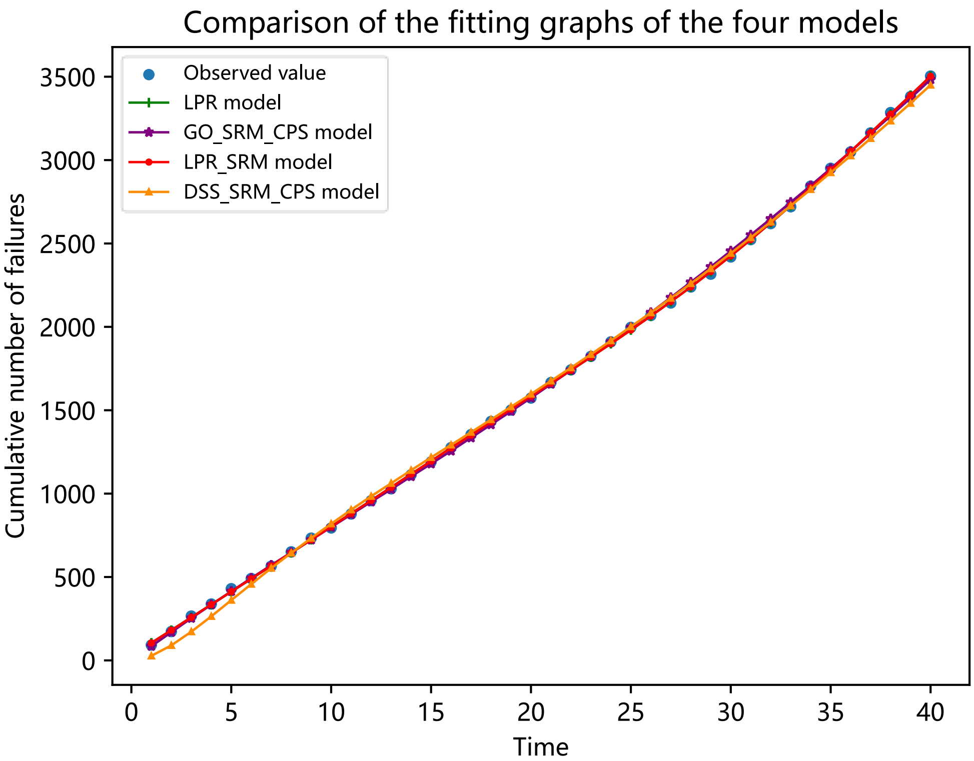

Figure 5. Comparison of the fitting graphs of the four models.

It can be seen from the above figure that the fitting effects of the four models are all relatively good. In the time period [0, 10], it can be clearly seen that the LPR_SRM model has the best fitting effect, followed by the LPR model, and the DSS_SRM_CPS model has the worst fitting effect.

5. Conclusions and discussions

5.1. Summary

This paper proposed an additive reliability model for the CPS based on local polynomial regression under masked data, namely the LPR_SRM model. Traditional parametric models require prior assumptions about the global functional form of the data, whereas the LPR_SRM model does not necessitate such assumptions. This enables the model to flexibly adapt to nonlinearities, heteroscedasticity, and local fluctuations in the data, thereby achieving precise characterization of complex data structures. Additionally, by integrating local fitting, kernel function weighting, and bandwidth adjustment, the LPR_SRM model maintains the high degree of freedom inherent in nonparametric methods while also considering computational efficiency and robustness. Based on the real software failure data containing masked data, the performance of the LPR_SRM model is compared with that of the LPR model, the GO model, the DSS model, the GO_SRM model and the DSS_SRM model. It is found that the LPR_SRM model has the best performance, followed by the LPR model. Based on the simulated cumulative failure data of the CPS containing masked data, by calculating the goodness-of-fit evaluation indicators of the LPR_SRM model, the LPR model, the GO_CPS model, the DSS_CPS model, the GO_SRM_CPS model and the DSS_SRM_CPS model, as well as drawing their respective fitting graphs for comparative analysis, it is known that the LPR_SRM model has the best fitting effect. To sum up, it is concluded that compared with the above models, the LPR_SRM model is more suitable for CPS reliability prediction under masked data.

5.2. Future Work

In the current frontier field of the reliability research on CPS, it is of crucial significance to explore the reliability of CPS based on masked data. The existence of masked data often has a significant impact on the accuracy and prediction effect of models. This paper has carried out a preliminary discussion on the CPS additive reliability model based on local polynomial regression under masked data. We are well aware that it is far from enough to merely limit the research to the existing types of masked data. Therefore, an in-depth and comprehensive exploration of the types of masked data is needed. One of the future research directions is to consider introducing the situation of general masked data, so as to further compare and analyze the performance of the CPS additive reliability model based on local polynomial regression proposed in this paper with that of traditional CPS reliability models. This improvement can provide more practical model selection and optimization schemes for the reliability analysis based on masked data.

Acknowledgement

This work was supported by Guangxi Natural Science Foundation (No. 2025GXNSFAA069686), Guangxi Science and Technology Base and Special Talents (No. Guike AD23023003), National Natural Science Foundation of China (No. 72361008), Science and Technology Plan Project of Guizhou Province (No.QianKeHeZhiCheng[2023]-General 268).

References

1. Xu J, Yu Q. Harmonic transform-based non-parametric density estimation method for forward uncertainty propagation and reliability analysis. Structural Safety, 2023, 103. https://doi.org/10.1016/j.strusafe.2023.102331
2. Li H-F, Lu M-Y, Zeng M, et al. A Non-Parametric Software Reliability Modeling Approach by Using Gene Expression Programming. Journal of Information Science and Engineering, 2012, 28(6): 1145-60.
3. Barghout M. A Two-Stage Non-Parametric Software Reliability Model. Communications in Statistics-Simulation and Computation, 2020, 49(5): 1159-80. https://doi.org/10.1080/03610918.2016.1189567
4. Wang Z, Wang J, Liang X J J O a S. Non-parametric estimation for NHPP software reliability models. Journal of Applied Statistics, 2007, 34(1): 107-19. https://doi.org/10.1080/02664760600994497
5. Gweon H, Lu J. A nearest neighbor-based approach for improving the reliability of multiclass probabilistic classifiers. International Journal of Data Science and Analytics, 2024. https://doi.org/10.1007/s41060-024-00624-8
6. Yu Q, Xu J. Distribution reconstruction and reliability assessment of complex LSFs via an adaptive Non-parametric Density Estimation Method. Reliability Engineering & System Safety, 2025, 254. https://doi.org/10.1016/j.ress.2024.110609
7. Roy A, Manna R, Chakraborty S J P E M. Support vector regression based metamodeling for structural reliability analysis. Probabilistic Engineering Mechanics, 2019, 55: 78-89. https://doi.org/10.1016/j.probengmech.2018.11.001
8. Hu B, Li Y, Yang H, et al. Wind speed model based on kernel density estimation and its application in reliability assessment of generating systems. Journal of Modern Power Systems and Clean Energy, 2017, 5(2): 220-7. https://doi.org/10.1007/s40565-015-0172-5
9. Cai J, Shi Y, Yue H. Accelerated life tests for log-normal series system with dependent masked data under Type-I progressive hybrid censoring. Communications in Statistics-Simulation and Computation, 2017, 46(2): 1628-46. https://doi.org/10.1080/03610918.2015.1045078
10. Liu B, Shi Y, Cai J, et al. Reliability analysis of masked data in adaptive step-stress partially accelerated lifetime tests with progressive removal. Communications in Statistics-Theory and Methods, 2017, 46(12): 6174-91. https://doi.org/10.1080/03610926.2015.1122058
11. Sarhan A M, El-Bassiouny A H J a M, Computation. Estimation of components reliability in a parallel system using masked system life data. Applied Mathematics and Computation, 2003, 138(1): 61-75. https://doi.org/10.1016/S0096-3003(02)00099-1
12. Zhao B, Yang J, Zhao M, et al. Wireless sensor network reliability modelling based on masked data. International Journal of Sensor Networks, 2015, 17(4): 217-23. https://doi.org/10.1504/IJSNET.2015.069584
13. Sarhan A M, Guess F M, Usher J S J I T O R. Estimators for reliability measures in geometric distribution model using dependent masked system life test data. IEEE Transactions on Reliability, 2007, 56(2): 312-20. https://doi.org/10.1109/TR.2007.895300
14. Sarhan A M J R E, Safety S. Estimations of parameters in Pareto reliability model in the presence of masked data. Reliability Engineering & System Safety, 2003, 82(1): 75-83. https://doi.org/10.1016/S0951-8320(03)00125-X
15. Yang J, Zhao M, Chen J. ELS algorithm for estimating open source software reliability with masked data considering both fault detection and correction processes. Communications in Statistics-Theory and Methods, 2022, 51(19): 6792-817. https://doi.org/10.1080/03610926.2020.1866610
16. Zheng Z, Yang J, Huang J. Software-hardware embedded system reliability modeling with failure dependency and masked data. Computers & Industrial Engineering, 2023, 186. https://doi.org/10.1016/j.cie.2023.109746
17. Liu B, Shi Y, Ng H K T, et al. Nonparametric Bayesian reliability analysis of masked data with dependent competing risks. Reliability Engineering & System Safety, 2021, 210. https://doi.org/10.1016/j.ress.2021.107502
18. Kuo L, Yang T Y J S, Letters P. Bayesian reliability modeling for masked system lifetime data. Statistics & Probability Letters, 2000, 47(3): 229-41. https://doi.org/10.1016/S0167-7152(99)00160-1
19. Sarhan A M J R E, Safety S. Reliability estimations of components from masked system life data. Reliability Engineering & System Safety, 2001, 74(1): 107-13. https://doi.org/10.1016/S0951-8320(01)00072-2
20. Srivastava P W, Agarwal M L. Masked Data Analyses of Reliability Systems Under Accelerated Life Testing [M]//KARANKI D R. Frontiers of Performability Engineering: In Honor of Prof KB Misra. Singapore; Springer Nature Singapore. 2024: 103-38. https://doi.org/10.1007/978-981-99-8258-5_5

21. Wang Y, Tong J, Pohl E A, et al. A Grey Systems Model for Estimating Component Reliability from Masked System Life Data. Journal of Grey System, 2013, 25(1): 96-109.
22. Sarhan A M J a M, Computation. Estimation of system components reliabilities using masked data. Applied Mathematics and Computation, 2003, 136(1): 79-92. https://doi.org/10.1016/S0096-3003(02)00017-6
23. Alemayehu T S, Kim J-H. Dependability analysis of cyber physical systems. Iet Computers and Digital Techniques, 2017, 11(6): 231-6. https://doi.org/10.1049/iet-cdt.2016.0164
24. Andrade E, Nogueira B, Callou G, et al. Dependability analysis of a cyber-physical system for smart environments. Concurrency and Computation-Practice & Experience, 2019, 31(1). https://doi.org/10.1002/cpe.4739
25. Lawal O A, Teh J. A framework for modelling the reliability of dynamic line rating operations in a cyber-physical power system network. Sustainable Energy Grids & Networks, 2023, 35. https://doi.org/10.1016/j.segan.2023.101140
26. Yang Y, Wang S, Wen M, et al. Reliability modeling and evaluation of cyber-physical system (CPS) considering communication failures. Journal of the Franklin Institute-Engineering and Applied Mathematics, 2021, 358(1): 1-16. https://doi.org/10.1016/j.jfranklin.2018.09.025
27. Gholami M, Gholami A, Mohammadtaheri M. Cyber- physical power system reliability assessment considering multi-state independent components. Electric Power Systems Research, 2023, 217. https://doi.org/10.1016/j.epsr.2023.109141
28. He R, Xie H, Deng J, et al. Reliability Modeling and Assessment of Cyber Space in Cyber-Physical Power Systems. Ieee Transactions on Smart Grid, 2020, 11(5): 3763-73. https://doi.org/10.1109/TSG.2020.2982566
29. Liu W, Gong Q, Han H, et al. Reliability Modeling and Evaluation of Active Cyber Physical Distribution System. Ieee Transactions on Power Systems, 2018, 33(6): 7096-108. https://doi.org/10.1109/TPWRS.2018.2854642
30. Gong H, Li R, An J, et al. Reliability Modeling and Assessment for a Cyber-Physical System With a Complex Boundary Behavior. Ieee Transactions on Reliability, 2023, 72(1): 224-39. https://doi.org/10.1109/TR.2022.3160460
31. Rostami A, Mohammadi M, Karimipour H. Reliability Assessment of Cyber-Physical Generation System. Iranian Journal of Science and Technology-Transactions of Electrical Engineering, 2023, 47(2): 617-26. https://doi.org/10.1007/s40998-022-00566-6
32. Zhou J, Li L, Vajdi A, et al. Temperature-Constrained Reliability Optimization of Industrial Cyber-Physical Systems Using Machine Learning and Feedback Control. Ieee Transactions on Automation Science and Engineering, 2023, 20(1): 20-31. https://doi.org/10.1109/TASE.2021.3062408
33. Ju Y, Yang M, Chakraborty C, et al. Reliability-Security Tradeoff Analysis in mmWave Ad Hoc-based CPS. Acm Transactions on Sensor Networks, 2024, 20(2). https://doi.org/10.1145/3582556
34. Zhou Y, Yuan X, Wang Z, et al. Analysis of interlocking faults and reliability evaluation methods for communication networks of power information physical systems. Journal of Physics: Conference Series, 2024, 2703(1): 012051. https://doi.org/10.1088/1742-6596/2703/1/012051
35. Huang J, Yang J, Zheng Z, et al. A novel probabilistic modeling for multilateral random attacks in cyber-physical system reliability analysis. Quality and Reliability Engineering International, 2024, 40(5): 2620-37. https://doi.org/10.1002/qre.3533
36. 36. [36] Yang W H, Xu C, Pan M X, et al. Ensure: Towards Reliable Control of Cyber-Physical Systems Under Uncertainty. Ieee Transactions on Reliability, 2023, 72(1): 289-301. https://doi.org/10.1109/TR.2022.3167116
37. Gong H F, Li R F, An J Y, et al. Reliability Modeling and Assessment for a Cyber-Physical System With a Complex Boundary Behavior. Ieee Transactions on Reliability, 2023, 72(1): 224-39. https://doi.org/10.1109/TR.2022.3160460
38. Lim T J R E, Safety S. Estimating system reliability with fully masked data under Brown-Proschan imperfect repair model. Reliability Engineering & System Safety, 1998, 59(3): 277-89. https://doi.org/10.1016/S0951-8320(97)00085-9
39. Liu X, Xie N. Grey-based approach for estimating software reliability under nonhomogeneous Poisson process. Journal of Systems Engineering and Electronics, 2022, 33(2): 360-9. https://doi.org/10.23919/JSEE.2022.000038
40. Salibian-Barrera M. Robust nonparametric regression: Review and practical considerations. Econometrics and Statistics, 2023. https://doi.org/10.1016/j.ecosta.2023.04.004
41. Davila-Frias A, Yodo N, Yadav O P, et al. Probabilistic modeling of hardware and software interactions for system reliability assessment. Quality Engineering, 2024, 36(1): 131-49. https://doi.org/10.1080/08982112.2023.2274563
42. Doksum K, Peterson D, Samarov A. On Variable Bandwidth Selection in Local Polynomial Regression. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2002, 62(3): 431-48. https://doi.org/10.1111/1467-9868.00242
43. Wang F, Gu X, Sun J, et al. Learning-based local weighted least squares for algebraic multigrid method. Journal of Computational Physics, 2023, 493: 112437. https://doi.org/10.1016/j.jcp.2023.112437
44. Yang J, Ding M, He M, et al. SDE-based software reliability additive models with masked data using ELS algorithm. Journal of King Saud University-Computer and Information Sciences, 2024, 36(3). https://doi.org/10.1016/j.jksuci.2024.101978
45. Kuchibhotla A K, Patra R K. On Least Squares Estimation Under Heteroscedastic And Heavy-Tailed Errors. Annals of Statistics, 2022, 50(1): 277-302. https://doi.org/10.1214/21-AOS2105